Биометрия

Биометрия – наука о статистическом анализе групповых свойств в биологии. Эти свойства могут изучаться при любой численности.

Методы биометрии позволяют дать математически точные характеристики свойств и признаков совокупностей, выявить степень генетического разнообразия признака и влияния на него различных факторов, прогнозировать эффект селекции.

Генеральная совокупность – большой массив животных, интересующих исследователя. Объем генеральной совокупности может быть как очень большим, так и очень малым.

Метод сплошного обследования, т.е. изучения всех членов генеральной совокупности, требует большой затраты времени и труда. Для изучения генеральной совокупности составляют выборку.

Выборка должна быть типичной, т.е. правильно отражать генеральную совокупность. Например, при изучении молочной продуктивности коров нельзя включать в выборку больных животных и животных с атрофией сосков, т.к. они нетипичны для изучаемой совокупности. Выборка должна быть однородной (одна порода, один пол). Она составляется по принципу случайного отбора.

При изучении генеральной совокупности по выборке неизбежны ошибки репрезентативности, указывающие на степень соответствия выборочных показателей параметрам генеральной совокупности.

Величина признака у отдельной особи называется вариантой и обозначается буквой x (икс), число особей обозначается буквой n.

К числу важнейших показателей, используемых в генетике и селекции, относятся следующие:

1. Средние величины средняя арифметическая (M) и средняя взвешенная (M взв.)
2. Показатели разнообразия признака: лимиты, среднее квадратическое отклонение, коэффициент вариации, нормированное отклонение;
3. Показатели связи между признаками: коэффициенты фенотипической и генетической корреляции, регрессии;
4. Показатели достоверности: ошибки репрезентативности, критерии достоверности;
5. Показатели наследуемости: коэффициент наследуемости и повторяемости.

Средняя величина вычисляется прямым способом, который заключается в суммировании всех вариант с последующим делением суммы на число вариант в выборке.

Средняя взвешенная представляет собой результат усреднения средних арифметических нескольких совокупностей. Например, известны средняя живая масса и число коров в трех хозяйствах: в первом средняя живая масса равна 420 кг, число животных 1000, во втором – 460 кг и 500 коров, в третьем 520 кг и 2000 коров. Средняя живая масса по данным трех хозяйств составила:

M взв = (420 x 1000 + 460 x 500 + 520 x 2000) / (1000 + 500 + 2000).

Показателем разнообразия в совокупности служат лимиты, которые характеризуют минимальное и максимальное значении изучаемого признака в выборке. Однако эти показатели недостаточны, т.к. животные с такими показателями могут быть нехарактерными для данной совокупности. Кроме того, лимиты не отражают индивидуальных различий внутри выборки. При одной и той же средней величине выборки могут быть неоднородны.

Наилучшим показателем разнообразия признака является среднее квадратическое отклонение (сигма), которое учитывает отклонение каждой варианты от средней арифметической. Вычисляют вначале среднюю арифметическую (М), затем находят отклонение (разность) каждой варианты от средней арифметической (х – М). Полученные отклонения суммируют с учетом знаков. При правильно вычисленной средней арифметической сумма должна равняться нулю. Затем каждой полученное отклонение возводят в квадрат и получают сумму квадратов. Квадраты отклонений всегда положительны. Полученное число сумм квадратов отклонений используют для вычисления сигмы. Формула приведена в практикумах, учебниках.

Среднее квадратическое отклонение величина именованная. При изучении разноименных признаков, разных по величине групп, сравнивать изменчивости невозможно. В таких случаях используют коэффициент вариации, который выражен в процентах.

При изучении совокупностей животных по нескольким признакам обнаруживается, что между ними существуют взаимосвязи. Взаимная связь между признаками в их изменении называется корреляцией. По форме корреляция может быть прямолинейной и криволинейной, по направлению прямой (положительной) и обратной (отрицательной).

Степень связи между признаками измеряется при помощи коэффициентов корреляции и регрессии. Коэффициент корреляции показывает направление связи и ее величину.

При положительной связи с увеличением (уменьшением) одного признака другой также увеличивается (уменьшается)

При отрицательной связи при увеличении одного признака, другой связанный с ним уменьшается.

Связь считается большой если коэффициент корреляции больше или равняется 0,7; средней – если коэффициент оказывается в пределах 0,4 – 0,6; маленькой – если коэффициент меньше 0,3. Таким образом величина коэффициента корреляции находится в пределах от -1 до +1. Величина эта отвлеченная измеряется в долях единицы.

Коэффициент же регрессии показывает не только направление связи, но и показывает, на сколько увеличивается или уменьшается первый признак при изменении второго на 1 и, наоборот, на сколько изменяется второй признак при изменении первого на 1. Полученные величины при этом имеют единицы измерения.

При характеристике генеральной совокупности сделанной по результатам расчетов выборки допускают наличие ошибок. Эти ошибки называются

ошибками репрезентативности. Величину ошибки репрезентативности определяют только для выборочных показателей.

Величина ошибки репрезентативности зависит от показателя изменчивости (величины сигмы) и от объема выборки (n).

При сравнении средних арифметических двух групп животных определяют достоверность разности двух средних величин (td).

Через коэффициент корреляции или коэффициент регрессии, рассчитанные по данным пар мать – дочь можно рассчитать коэффициент наследуемости. Коэффициент наследуемости в свою очередь используется при прогнозировании будущей продуктивности (эффект селекции).

Обработка экспериментальных данных

Обработку полученных данных проводят биометрически с определением средней величины, ее ошибки, достоверности разности, корреляции, наследуемости и т.д. Вычисление этих величин позволяет делать по результатам исследований объективные и достоверные выводы. Работа с данными, необработанными биометрически, к защите не допускаются.

В настоящие время для биометрической обработки широко используются персональные компьютеры, которые оснащены пакетом программ для статистического анализа. Проведение биометрической обработки данных обычными методами довольно трудоемкая работа и отнимает у исследователя много времени. Использование для этой цели компьютеров существенно ускоряет решение биометрических задач.

На кафедре генетики и селекции разработаны методические указания по биометрической обработке экспериментальных данных на компьютере с использованием средств электронных таблиц Excel, которые рекомендуются для применения при выполнении выпускных квалификационных работ.

Средние величины и показатели вариации признака являются основными показателями для характеристики совокупности. К первым относится средняя арифметическая. Ко вторым – стандартное отклонение и коэффициент вариации.

Средняя арифметическая (M) показывает, какое значение признака наиболее характерно в целом для данной совокупности. Рассчитывается по формуле:

$$M = \frac{\sum x_n}{n}$$

Ошибку репрезентативности средней арифметической (m) определяют по формуле:

$$m = \frac{\sigma}{\sqrt{n}}$$

Стандартное отклонение или среднее квадратическое отклонение используют для характеристики изменчивости признака. Стандартное отклонение обозначается буквой σ (сигма). Рассчитывается по формуле:

$$\sigma = \sqrt{\frac{\sum (x - M)^2}{n}}$$

Коэффициент вариации (C_v) используют для сравнения изменчивости признаков, выраженных в разных единицах, поэтому эта величина измеряется в процентах. Для расчета используют формулу:

$$C_V = \frac{\sigma}{M} \cdot 100\%$$

Коэффициент корреляции (r). Признаки и свойства животных находятся в определенной взаимосвязи. Для оценки силы и характера взаимосвязи между признаками вычисляют коэффициент корреляции.

$$r = \frac{\sum X + Y - \frac{\sum X \times \sum Y}{n}}{\sqrt{C_x \times C_y}},$$

где:

$$C_x = \sum X^2 - \frac{(\sum X)^2}{n}$$

$$C_Y = \sum X^2 - \frac{(\sum Y)^2}{n}$$

Ошибку репрезентативности коэффициента корреляции определяют по формуле:

$$m = \frac{1 - r^2}{\sqrt{n}}$$

Критерий достоверности коэффициента корреляции находим по формуле:

$$t = r/m$$

Критерий достоверности разности двух средних величин определяют по формуле:

$$t_d = \frac{M_1 - M_2}{\sqrt{m_1 + m_2}}$$

Затем находим *достоверность (P)* полученных результатов по Стьюденту.

*Коэффициент наследуемости (h^2)*определяют по формуле:

$$h^2 = 2 \text{ x } r$$

Обязательным условием работы со статистическим пакетом компьютерной программы является вызов Excel 7.

А) Процедура вызова Microsoft Excel 7

1. Установить курсор мыши на кнопку ПУСК и щелкнуть ее левой клавишей мыши.
2. В появившемся меню подать команду ПРОГРАММЫ.
3. В новом меню подать команду Microsoft Excel.

В результате определенных операций на экране появиться окно Microsoft Excel 7.

Б) Работа с книгой, листами и ячейками.

Как в типовом окне Windows, главное окно Excel 7 имеет заголовок, главное меню, панель инструментов, так называемую "Строку формул", сопроводительное окно и область "Рабочей книги".

Как и в обычной жизни, книга содержит листы, по умолчанию таких листов 3. Пользователь может добавить листы, но не более 255. Переход на листы осуществляется кликированием ярлычков, расположенных внизу книги. Листы книги можно перелистывать последовательно, для этой цели служит "Кнопка прокрутки ярлычков", расположенная в самом начале списка ярлычков.

По умолчанию книга носит название Книга1, но за пользователем всегда остается право дать книге свое название.

Ячейка - опорный элемент Электронных таблиц (ЭТ), практически все действия в той или иной степени связаны с ячейками ЭТ. К каждой ячейки можно адресоваться, подобно тому, как это делается на шахматной доске.

Столбцы таблицы названы буквами латинского алфавита, а строки - порядковыми числами начиная с 1. К любому данному можно обратиться по его уникальному адресу, который составляется из буквы- имени столбца и числа - имени строки. Например, A1 - адрес верхней левой ячейки таблицы.

Jn: Yv - ссылка на все ячейки, расположенные внутри прямоугольной области, которая определяется адреса в противоположных углах.

Текущая активная ячейка выделяется четырехугольной рамкой, которую называют " Указатель ячейки". После запуска программы Указатель ячейки, как правило, находится по адресу A1. Перемещать указатель можно либо стрелкой клавиатуры, либо указывать курсором мыши на ту ячейку, куда следует переместить Указатель ячейки. При этом следует щелкнуть левой клавишей мыши.

В активное окно можно внести любой текст, любые числа, а также вычислительные выражения.

Вся информация по текущей ячейки отображается в Строке формул, которая разделена на три части:

Первая часть отражает содержимое текущей ячейки. Средняя часть может быть либо пустой, либо содержать три кнопки:

Крестик- отмена последнего действия (дублирует клавишу Esc);

Галочка - соответствует клавише Enter.

f_x - вызов мастера функций.

Левая часть содержит адреса текущей ячейки.

Рассмотрим заполнение ячеек таблицы на конкретном примере.

Вначале необходимо ввести исходные данные. После введения числа в ячейку А1 необходимо нажать Enter чтобы перевести курсор в следующую нижнюю ячейку А2 или курсор можно перемещать стрелками на клавиатуре вправо (⟵), влево (⟶), вверх (↑), вниз (↓). В столбце А показаны значения удоя за 350 дней лактации у 10 коров и столбец В показывает значения содержания жира в молоке этих коров.

После введения данных, активируем любую пустую ячейку. Если вы ввели большое количество признаков (столбцов), то удобнее активировать ячейку в конце столбца (снизу). В нашем примере это будет ячейка А11.

Среднее арифметическое значение можно рассчитывать двумя способами:

1. Вызвать мастера функций, щелкнув левой клавишей мыши на панели инструментов значок *fx*. После этого появиться окно, показывающее слева КАТЕГОРИИ (10 недавно использовавшихся, полный алфавитный перечень, финансовые и т.д.). Необходимо выбрать в категории СТАТИСТИЧЕСКИЕ, щелкнув левой клавишей мыши и снизу или справа, будут показаны ФУНКЦИИ в алфавитном порядке. Необходимо найти СРЗНАЧ, затем либо сначала нажать левой клавишей мыши и затем на кнопку ОК, либо два раза на СРЗНАЧ.

Формирование экспериментальных данных (пример)

	A	B
1	4272	3,83
2	3859	3,76
3	4584	3,85
4	4807	3,8
5	4774	4,04
6	4872	4,02
7	4006	3,72
8	5539	4,09
9	5110	3,77
10	5180	4,02

После этого появиться окно, где необходимо ввести диапазон данных для расчета среднего значения в верхнюю строку называемой ЧИСЛО 1 (ее можно в последующем в любой момент исправить). В нашем примере это будет диапазон А1:А10. Затем нажать на ОК или Enter. Верхнем окне появится формула с введенным диапазоном расчета, в ячейке А11 будет показано среднее значение, в данном примере будет равно 4700,3 кг. Аналогичным путем можно рассчитать среднее значение и для содержания жира в молоке. Для этого надо выбрать другую ячейку, например, В11. Среднее значение содержания жира составляет в данном примере 3,89%.

2. Также можно ввести формулу без вызова мастера функции набрав в активной ячейке = СРЗНАЧ (А1:А10) и затем нажав клавишу Enter.

Для расчета стандартного отклонения необходимо активировать вторую ячейку, например А12. Стандартное отклонение также можно рассчитать двумя способами: вызовом мастера функции: выбрав категорию Статистические и функцию СТАНДОТКЛОН, затем нажать ОК и обозначить диапазон данных расчета также как и для среднего арифметического (А1:А10). Либо можно набрать в активной ячейки = СТАНДОТКЛОН (А1:А10) и затем, нажав на клавишу Enter. Стандартное отклонение по удою составляет в данном примере ±530,67 кг, а по содержанию жира ± 0,14%.

Для расчета ошибки репрезентативности следует активировать следующую ячейку А13 и набрать =А12/КОРЕНЬ(СЧЕТ(А1:А10)) (в данном примере А12 это стандартное отклонение) и затем, нажав на клавишу Enter, в этой ячейке появиться значение ошибок репрезентативности средней арифметической по удою за 305 дней лактации, которая в данном примере составляет ± 167,81кг; по содержанию жира ± 0,04%.

Коэффициент вариации - это отношение среднеквадратического отклонения к арифметической величине, выраженное в процентах. Рассчитаем коэффициент вариации в ячейке А14, где набираем следующую формулу =А12/А11х100, а затем нажимаем на клавишу Enter и в ячейке А14 появится значение коэффициента вариации. В данном примере коэффициент вариации удоя равен 11,29%, а содержания жира - 3,53%.

Для расчета коэффициента корреляции можно вызвать мастера функций, предварительно активировав, пустую ячейку (например, А16), выбрав в категориях СТАТИСТИЧЕСКИЕ найти в функциях КОРРЕЛ и нажать ОК. Затем появиться окно, в котором следует указать в МАССИВЕ1 диапазон данных первого признака, в МАССИВЕ 2 диапазон данных второго признака и нажать на ОК. В верхнем окне увидим следующую формулу =КОРРЕЛ (А1:А10;В1:В10), а в ячейке А16 появится результат. Также можно просто набрать формулу = КОРРЕЛ (А1:А10;В1:В10) и нажав на клавишу Enter получим результат. Если вы решили считать не под столбцами (например в столбце С и т.д.), то можно указать только столбцы (А:А;В:В), это же касается и предыдущих форм и расчетов. В данном примере значение коэффициента корреляции между удоем и содержанием жира составляет - 0,705.

Для расчета ошибки репрезентативности коэффициента корреляции активируем следующую ячейку А17 и набрав формулу =(1-А16^2)/(КОРЕНЬ(СЧЕТ(А1:А10))) и затем, также нажав на клавишу Enter, получим результат. В данном примере она равна 0,159.

Критерий достоверности можно найти, введя в активированную (например, А18) =А16/А17 и затем, необходимо нажать клавишу Enter , в результате получим значение критерия достоверности (в данном примере 4,43),

после которого найдем вероятность, сравнивая значение со значениями различных порогов. Формула будет выглядеть следующим образом, которую необходимо ввести в активированную ячейку, (например, А19)=ЕСЛИ(А18<1,96;то"p<0,05",ЕСЛИ(А18>2,57;ЕСЛИ(А18<3,3;"p<0,01","p<0,001")) и затем, необходимо нажать клавишу Enter.

Для расчета коэффициента наследуемости необходимо вначале найти коэффициент корреляции между одноименными признаками у родственников. Например, удой за 305 дней лактации (кг) дочерей и удой за 305 дней лактации (кг) у матерей. В свободную ячейку необходимо ввести адрес (номер ячейки, в которой показан коэффициент корреляции) и умножить на 2. При этом формула будет выглядеть следующим образом: =B19x2 и в этой ячейке появится результат.

Критерий достоверности разности двух средних величин можно найти =ПРОИЗВЕД (СУММ(A11;-D11); 1/КОРЕНЬ (СУММ(A13*A13;D13*D13))),
A11 - средняя арифметическая удоя коров за 305 дней лактации, A13 ошибка репрезентативности первого признака, и если ввести среднее арифметическое значение удоя за 305 дней лактации матерей этих коров в ячейку с адресом D11, а в D13 - ошибку репрезентативности средней арифметической удоя за 305 дней лактации матерей коров за 305 дней лактации.

Если Вы работаете с большим массивом, то для удобства можно использовать ФИЛЬТР, который может разделить количественные значения по диапазонам, которые Вы сами обозначите. Для этого необходимо вначале выделить столбец, затем на верхней панели инструментов найти ДАННЫЕ, щелкнув два раза левой клавишей мыши открыть окно, в котором Вы можете найти ФИЛЬТР и затем выбрать АВТОФИЛЬТР и щелкнуть два раза левой клавишей мыши, исходные данные будут отсортированы.